



Porter, D., Basu, N. and Siebert, S. 2020) Classification criteria: time for a rethink? *Annals of the Rheumatic Diseases*, (doi: 10.1136/annrheumdis-2020-217769).

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/218869/>

Deposited on: 21 August 2020

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>

## **Classification criteria – time for a re-think?**

Duncan Porter<sup>1</sup>, Neil Basu<sup>2</sup> and Stefan Siebert<sup>2</sup>

<sup>1</sup> Corresponding author  
Gartnavel General Hospital  
1053 Great Western Road  
Glasgow  
G12 0YN  
United Kingdom  
duncan.porter@glasgow.ac.uk  
ORCID <https://orcid.org/0000-0001-8481-5204>

<sup>2</sup> Institute of Infection Immunity & Inflammation,  
Sir Graeme Davies Building,  
120 University Place  
Glasgow  
G12 8QQ  
United Kingdom

Word count 1756

The last decade has seen the development or renewal of classification criteria in sundry rheumatic diseases, including systemic lupus erythematosus (SLE),[1] rheumatoid arthritis,[2] and axial spondyloarthritis (axSpA).[3] These criteria seek the laudable aim of standardising the populations included in clinical trials and observational cohorts for research purposes. But to what extent have the benefits of classification criteria been realised? Have there been unintended consequences as their profile has grown? And could we better use criteria to achieve the desired end of facilitating the implementation and interpretation of research findings to enable their translation into clinical practice?

We are all familiar with the refrain of key opinion leaders when they present their update on the management of a disease – “These are classification criteria, not diagnostic criteria ...” But how often have we heard the same speaker move smoothly on to state, “... but I find them helpful in clinical practice too”? Indeed, the abstract of the original paper describing the validation and final selection of the classification criteria for axSpA concluded with the statement that “The new ASAS classification criteria for axial SpA ... may help rheumatologists in clinical practice in diagnosing axial SpA in those with chronic back pain.” [3]

However, the dangers of applying classification criteria to clinical practice for the purpose of diagnosis are easily demonstrated. The criteria are invariably developed and validated in specialist centres where there is a high prior (pre-test) probability of the disease, and they are evaluated based on a clinical diagnosis of the disease in question made by experts in that condition, after alternative causes or explanations for the patient’s symptoms have been excluded. When the criteria are applied to a different population - for instance, in primary care - where there is a low prior (pre-test) probability of the disease, they result in a large number of false positive diagnoses. Take the example of axSpA: a systematic literature review and meta-analysis identified seven studies that examined the ASAS criteria’s performance, with a pooled sensitivity 82%, specificity 87% and positive likelihood ratio (LR+) of 6.2.[4] Applying these figures in a secondary care setting with a prior (pre-test) probability of 0.3 (30%), results in a posterior (post-test) probability of 0.72 (72%) of those who fulfil the criteria. However, if the criteria were applied in a low risk population, with a prior (pre-test) probability of 0.05 (5%), then 75% of the people fulfilling the criteria would be given a false positive diagnosis (figure 1).

This is not a theoretical concern that is without practical implications: we can see the dangers worked out in practice in the literature. For instance, the authors of a study of chronic low back pain in primary care began with the flawed premise that the new ASAS classification criteria ‘will improve early diagnosis’. They studied young patients with chronic low back pain and found that 86/363 (24%) patients fulfilled the ASAS classification criteria for axSpA. The authors concluded that chronic low back pain in young patients is ‘frequently caused by undiagnosed axial SpA’. Of note, only 17/86 (20%) of the identified cases were HLA-B27 positive: it is highly likely that most of those fulfilling the ASAS classification criteria were false positive diagnoses based on the inappropriate use of the classification criteria.[5]

The pitfalls of the inappropriate use of classification criteria for the purpose of diagnosis are well known; but have the putative benefits of their use in research been realized? Let us use axSpA as an example again: like most of the inflammatory rheumatic conditions, axSpA is a spectrum that ranges from early/mild to severe/late disease, such that only some patients with severe disease ultimately exhibit the classical radiographic features of ankylosing spondylitis. Restricting recruitment to pivotal Phase 3 studies to those who fulfilled the modified New York criteria for ankylosing spondylitis led to approval for TNFi therapy being restricted to this population. Recognition of the limitations of the modified New York criteria were part of the driving force behind the development of the ASAS classification criteria for axSpA. Additional studies of patients with non-radiographic axSpA were required by the FDA before approval was granted for these patients – but these are part of the spectrum of the same condition[6] and it makes no sense to require drug development programs to study different sub-populations based on their classification status, rather than their clinical diagnosis. The tail has started to wag the dog!

Moreover, clinicians are left with the conundrum of how to apply the results of clinical studies to their practice. Using the fulfillment of classification criteria as an inclusion criterion for a clinical study risks the unnecessary exclusion of patients from the study who in real life could benefit (e.g. by limiting participation to those who meet the modified New York criteria) or the inappropriate inclusion of many patients without the disease (e.g. if the ASAS criteria were used to identify suitable recruits in primary care populations). How are clinicians to decide whether the results of a study apply to their patient? In clinical practice, we (ought to) make decisions based on the patient's clinical diagnosis (based on history, examination, investigations, and the exclusion of alternative explanations), not on whether he/she fulfills the classification criteria. It follows that the most appropriate inclusion criterion for a clinical trial ought to be a clinical diagnosis of sufficient disease activity/severity to warrant the treatment being studied.

Updated EULAR/ACR classification criteria for SLE have recently been published. The criteria use weighted scores for different clinical, immunological and pathological features in several domains; the accrual of 10 or more points leads to a classification of SLE. This creates a dichotomy of those who do/do not meet the classification criteria, but diagnosis is not binary. Firstly, does it ensure the standardization of a population of patients with 'definite SLE'? We know that the disease is heterogeneous, and that ethnicity, type of organ involvement and immunology are all strongly associated with differing outcomes – patients with neuropsychiatric lupus are very different from patients with predominantly cutaneous disease, even if both fulfill the classification criteria. Secondly, does the dichotomy result in the exclusion of some patients with a clinical diagnosis of SLE from studies? Without doubt! This has led to the development of the concept of 'incomplete lupus' representing patients with mild or early disease, and those with a restricted number of clinical features. These patients also have significant morbidity that requires research and treatment but they are excluded from studies which use the fulfillment of classification criteria as an entry criterion.

Hence we see randomized controlled trials being designed to test treatment efficacy in patients defined as having ‘incomplete lupus’.[7]

Classification criteria for RA were updated in 2010. These have been broadly welcomed as an improvement (in terms of sensitivity and applicability) in early disease. However, their use in clinical trials has led to some distortion of the populations who have been studied. For example, it is known that patients with sero-positive RA have a worse long term prognosis than those with sero-negative disease; there is also plentiful evidence that higher disease activity over time, is associated with poorer outcomes. Yet when we look at patients enrolled in trials in early RA we may observe a paradox – patients who are sero-negative have higher disease activity (e.g. in ARCTIC, sero-negative patients had DAS28 = 3.9 at baseline, compared to DAS28 = 3.4 in sero-positive patients).[8] The reason may be quite simple – the presence of strong positive rheumatoid factor (RF) or anti-citrullinated protein antibodies (ACPA) contributes three points towards the six required to fulfill the classification criteria for ‘definite’ RA. Denied these three points, patients who are sero-negative must have higher disease activity if they are to fulfill the criteria for RA. Put another way, two patients who have identical clinical features and acute phase response may be classified differently according to their auto-antibody status, one thereby being eligible for a study and the other being excluded. Now it may be that ACPA positive disease is fundamentally different from sero-negative disease, but if that is true it would make more sense to include patients on the basis of their ACPA status rather than conflating this with the fulfillment of the classification criteria.

These examples demonstrate that the use of classification criteria in clinical practice to make a diagnosis is inappropriate. It risks over-diagnosis, over-referral and over-treatment in populations with a low prior (pre-test) probability of disease. In research, the use of the fulfillment of classification criteria as the grounds for entry to a clinical trial leads to the distortion of the study population and inappropriate restrictions on the label when drugs are licensed. How then can classification criteria be used most effectively? They should be used as a lens through which the study population can be viewed, such that different studies can be compared and contrasted. The main entry criterion for a clinical trial ought to be a clinician diagnosis with sufficient disease activity (or severity) to justify the planned treatment. At first sight, the proposal to revert to using a clinical diagnosis as an entry criterion to a clinical study risks a return to the ‘bad old days’ with the danger that investigators from different backgrounds with varying levels of expertise would recruit inappropriate patients into studies. This is to misunderstand the problem because, as we have shown, a high prior (pre-test) probability of disease is a *necessary pre-requisite* for the appropriate use of classification criteria. If the clinical diagnoses cannot be trusted, then the fulfilment of the classification criteria will not guarantee a high posterior probability of the disease either. As the authors of the ACR/EULAR SLE classification criteria point out, scoring is a process that requires “diligence and clinical experience.”[1] The robustness of the diagnoses on a population level will be derived from describing the cohort in detail, *including the proportion that fulfill the relevant classification criteria*. This has several advantages: (1) it will help to avoid imposing an artificial dichotomy on conditions that represent a continuous spectrum; (2) if necessary, a *priori* sub-

group analyses can be planned to compare those who fulfill the criteria with those who don't; (3) it will allow nuanced characterisation of research cohorts (e.g. in SLE, describing the average number of points scored, and in which domains).

Much useful work has been undertaken during the development of classification criteria, and we should not 'throw the baby out with the bath water'. But clinicians must resist the temptation to use the criteria in patient diagnosis, and the use of the criteria to define which patients are recruited into clinical trials needs to evolve.

### **Competing interests**

Neil Basu – Research grants: Vifor, GSK, Pfizer, Novartis. Advisory board: Lilly, MSD. Speaking fees: Roche, Vifor, Abbvie. Stefan Siebert – GRAPPA global steering committee, ASAS member; research grants, speaker or consultancy fees from AbbVie, BMS, Boehringer-Ingelheim, Celgene, GSK, Janssen, Novartis, Pfizer and UCB.

### **Patient and public involvement**

No patients were involved in the development of this article.

### **References**

- 1 Aringer M, Costenbader K, Daikh D, *et al.* 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. *Ann Rheum Dis* 2019;**78**:1151.
- 2 Aletaha D, Neogi T, Silman AJ, *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010;**69**:1580.
- 3 Rudwaleit M, van der Heijde D, Landewe R, *et al.* The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;**68**:777.
- 4 Sepriano A, Rubio R, Ramiro S, *et al.* Performance of the ASAS classification criteria for axial and peripheral spondyloarthritis: a systematic literature review and meta-analysis. *Ann Rheum Dis* 2017;**76**:886.
- 5 van Hooft L, Luime J, Han H, *et al.* Identifying Axial Spondyloarthritis in Dutch Primary Care Patients, Ages 20-45 Years, With Chronic Low Back Pain. *Arthritis Care Res (Hoboken)* 2014;**66**:446–53. doi:10.1136/annrheumdis-2012-201766

- 6 van der Heijde D, Ramiro S, Landewe R, *et al.* 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2017;**76**:978–91. doi:10.1136/annrheumdis-2016-210770
- 7 Olsen NJ, James JA, Arriens C, *et al.* Study of Anti-Malarials in Incomplete Lupus Erythematosus (SMILE): study protocol for a randomized controlled trial. *Trials* 2018;**19**:694.
- 8 Nordberg LB, Lillegraven S, Aga A-B, *et al.* Comparing the disease course of patients with seronegative and seropositive rheumatoid arthritis fulfilling the 2010 ACR/EULAR classification criteria in a treat-to-target setting: 2-year data from the ARCTIC trial. *RMD Open* 2018;**4**:e000752–2.

Figure 1 – nomogram to calculate the posterior (post-test) probability according to the prior (pre-test) probability and likelihood ratio. In axSpA, the ASAS classification criteria have sensitivity 82%, specificity 87%, and positive likelihood ratio (LR+) 6.2:

- A. Where the prior (pre-test) probability = 0.3, the posterior (post-test) probability = 0.72 in those who fulfill the criteria
- B. Where the prior (pre-test) probability = 0.05, the posterior (post-test) probability = 0.25 in patients who fulfill the criteria